* GET OUT NOTES/HW

* HW QUIZ 2

* CROSS OUT:
#1, 7

*** Inputting data into the calculator.**
Page 14 in the book has this info if you forget

*** Typing data into L1**
*** Naming a list**

Now do the following:
Person 1 --> Name a list SIBS and input the class data for the # of siblings people have

Person 2--> Name a list COLLG and input the class data for the number of colleges everyone is applying to. For N/A, put in 0.

*** Sorting data**

*** Transferring from one calculator to another:**
Transfer the list SIBS from person 1 to person 2

Transfer the list COLLG from person 2 to person 1

*** Transfer the following lists to your calculators:**

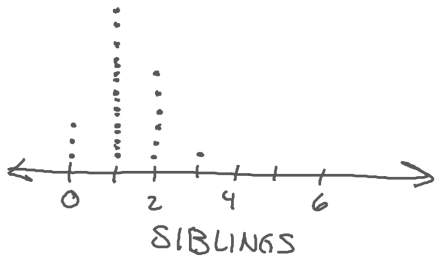
BLK1	INCOM	QUES1	SMOKE
BLK2	NJGAS	QUES4	STATE
BLK3	NONSM	SATMF	TEST
GPA	PAGAS	SATMM	TST1F
	PRES		TST1M

AP STAT: CHAPTER 4 & 5: QUANTITATIVE DATA

Quantitative Distributions:

1) Dotplots

Example: number of siblings of your classmates



2) Stemplot (aka Stem and Leaf Plot)

Example: Babe Ruth's homerun totals each season for the Yankees:
~~7,~~ 54, 59, 35, 41, 46, 25, 47, 60, 54, 46, 49, 41, 34, 22

Create a stemplot

Babe Ruth's HR totals

Key: 3|4 = 34 homeruns

1	
2	2 5
3	4 5
4	1 1 6 6 7 9
5	4 4 9
6	0

- no commas
- title
- key

Example: Babe Ruth's homerun totals each season for the Yankees:
54, 59, 35, 41, 46, 25, 47, 60, 54, 46, 49, 41, 34, 22

Roger Maris' homerun totals for the Yankees:
~~08~~, ~~13~~, ~~23~~, 33, ~~28~~, ~~16~~, ~~14~~, 39, ~~26~~, 61

3) Create a **Back-to-Back stemplot** with Babe Ruth's

Roger Maris HR totals

Babe Ruth's HR totals

Key: 3|4 = 34 HR's

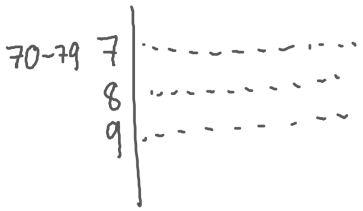
8	0	
6 4 3	1	
8 6 3	2	2 5
9 3	3	4 5
	4	1 1 6 6 7 9
	5	4 4 9
1	6	0

SPLITTING STEMS:

When? When your data is clumped together within only a few stems.

How can stems be split? By 5, 2, 1

AGE GUESS EXAMPLE



Example: Age guesses

Regular plot: By 10's

```
2 | 5 6 6 7 7 7 8 8 9 9 9 9 9
3 | 0 0 0 0 0 0 0 1 1 1 1 1 1 1 2 2 2 2 2 ...
4 | 0 1
```

By 5's

```
2 | 5 6 6 7 7 7 8 8 9 9 9 9 9
3 | 0 0 0 0 0 0 0 1 1 1 1 1 1 1 2 2 2 2 2  3 ...
3 | 5 5 6 6 7 7 8 8 9 9
4 | 0 1
```

By 2's

```
20,21 | 2 |
22,23 | 2 |
24,25 | 2 | 5
26,27 | 2 | 6 6 7 7 7
28,29 | 2 | 8 8 9 9 9 9 9
      | 3 | 0 0 0 0 0 0 0 1 1 1 1 1 1 1
      | 3 | 2 2 2 2 2 3 3 3 3 3 3 3
      | 3 | 4 4 4 4 4 4 4 4 5 5
      | 3 | 6 6 7 7
      | 3 | 8 8 9 9
      | 4 | 0 1
```

By 1's

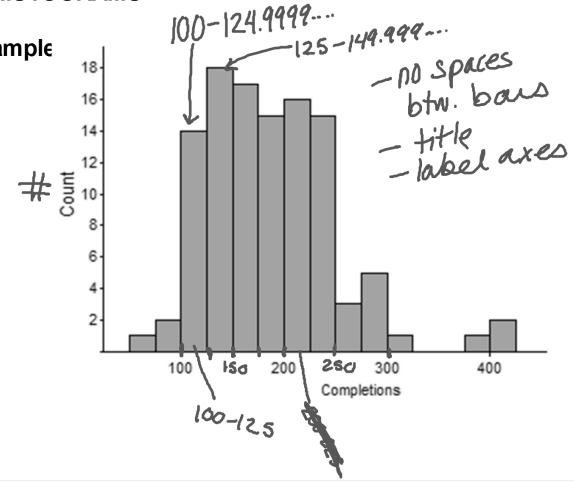
```
2 |
2 |
2 | 5
2 | 6 6
2 | 7 7 7
2 | 8 8
2 | 9 9 9 9 9
3 | 0 0 0 0 0 0 0
3 | 1 1 1 1 1 1 1
3 | 2 2 2 2 2
3 | 3 3 3 3 3 3 3
3 | 4 4 4 4 4 4 4 4
3 | 5 5
3 | 6 6
3 | 7 7
3 | 8 8
3 | 9 9
4 | 0
4 | 1
```

BLOCK 2 BLOCK 3

```
                    |2|
                    |2|
                    |2| 5
              7 7   |2| 6 6 7
        9 9 9 8 8   |2| 9 9
1 1 1 1 0 0 0 0 0   |3| 0 0 1 1 1
        3 3 3 3 2   |3| 2 2 2 2 3 3 3
          5 4 4 4   |3| 4 4 4 4 4 5
              6 6   |3| 7 7
              9 8   |3| 8 9
                0   |4|
                    |4| 1
```

4) HISTOGRAMS

Example



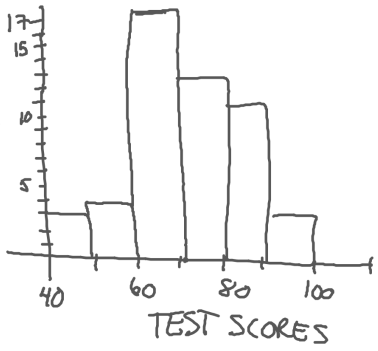
ple: The following are a list of test scores on an exam. Create a histogram of these scores.

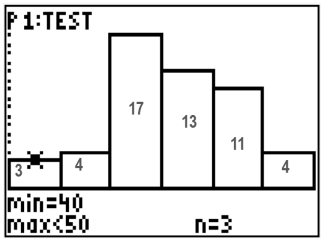
61	66	68	74	77	84	91	76	82	87	53	64	57	51	67	81	65
62	66	69	75	78	84	95	76	82	90	67	71	59	64	70	85	68
64	66	69	76	80	85	96	64	68	72	76	81	86	99	76	96	72

NS =

** USE LIST TEST

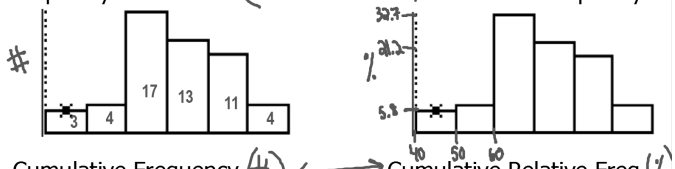
-min. of 5 bars





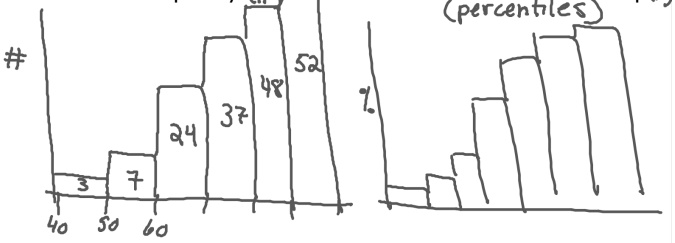
4 types of histograms:

Frequency ↔ Relative Frequency



Cumulative Frequency (#) ↔ Cumulative Relative Freq (%)

(percentiles)



Example:

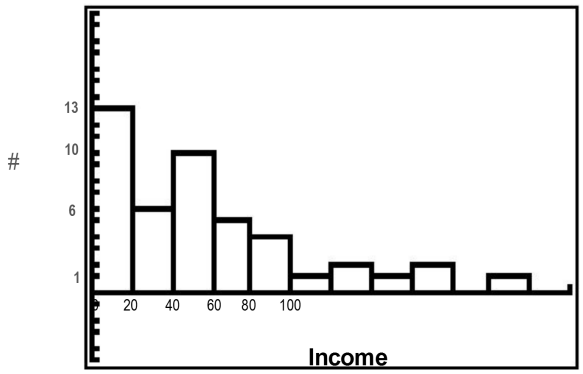
Using the list INCOM, create a frequency histogram

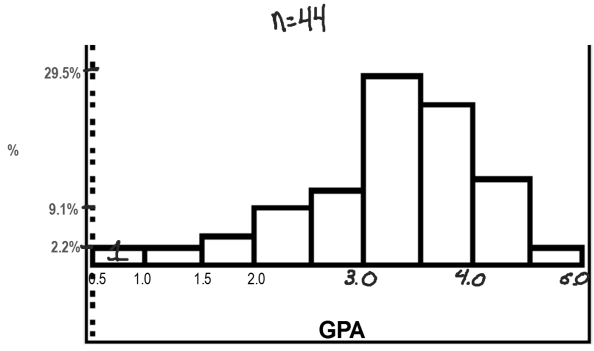
data: 3 = $3,000

Example:

Using the list GPA, create a relative frequency histogram

Answers to examples:





EXTRA EXAMPLE:

Create a frequency histogram from the list NJGAS
(these are prices of regular gas in NJ, in dollars)

Describing Distributions... numerically

CENTER: Median: middle obsevation,
50% above/below

Quartiles: medians of the lower & upper half of the data

Q3 = upper half Q1 = lower half

SPREAD: Range: (min, max)

IQR: Q3 – Q1= middle 50% of data

5# Summary: min, Q1, Med, Q3, max

CENTER: Mean: Arithmetic average

Formula:

$$\bar{x} = \frac{\sum x_i}{n}$$

sigma = sum

x_i ← # of piece of data

n ← total # of pieces of data

* Read as "x-bar"

SPREAD: Standard Deviation & Range

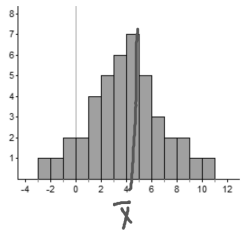
(a, b)

Standard Deviation (s) is

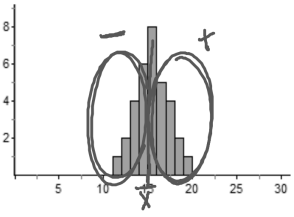
* number that describes how spread out the data is
* Average difference of each point from the mean.
* Higher the number --> more spread out the data

Examples:

large standard dev.



small standard dev.



Standard Deviation Formula:

$$s = \sqrt{\frac{\sum_{i=1}^{n}(y_i - \bar{y})^2}{n-1}}$$

On calculator...

1-var stats:

```
1-Var Stats
 x̄=9.88
 Σx=247
 Σx²=3839
 Sx=7.633915553
 σx=7.479679138
↓n=25
```

```
1-Var Stats
↑n=25
 minX=2
 Q1=5.5
 Med=7
 Q3=11.5
 maxX=35
```

Properties of the standard deviation (s):

* s = 0 when all data is the same

* Unless all points are the same, "s" is always positive

$\bar{x} = 40$ (40, 95)

* It is not resistant- It is affected by outliers

* variance = (std. dev.)^2

$\bar{x} = 72$

s^2 = variance = 16

$s = 4$

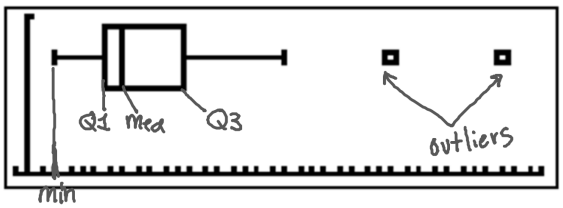
Summary:

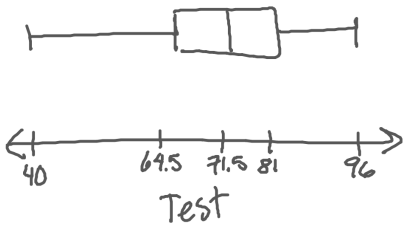
CENTER:	MEDIAN	MEAN
SPREAD:	RANGE IQR	STD. DEVIATION RANGE

Another Distribution:

Modified Boxplot

* Graphical Display of 5# summary
* If outliers are present the "whisker" extends to the last value not an outlier
* The outlier is marked with an x, box, or point





Formally determining outliers...

1.5 * IQR Rule

- Calculate the LOWER FENCE and the UPPER FENCE:
  - LF = Q_1 – 1.5(IQR)
  - UF = Q_3 + 1.5(IQR)
- Normal Data is inside (LF, UF). Outliers are anything outside that range.

Example: Using list STATE, test for outliers

IQR = 11.5 − 5.5 = 6

LF = 5.5 − (1.5 × 6) = −3.5

UF = 11.5 + (1.5 × 6) = 20.5

(−3.5, 20.5)

outliers @ 27 and 35

Example:

Using the following AGES in months data, test to see if there are any outliers

n = 52
min = 204
Q1 = 208.5
Med = 214
Q3 = 216
Max = 240

IQR = 7.5

LF = 208.5 − (1.5 × 7.5) = 197.25

UF = 216 + (1.5 × 7.5) = 227.25

(197.25, 227.25)

at least 1 @ 240 outliers??

Parallel boxplots:

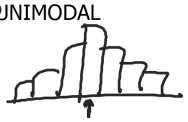
- 2 or more boxplots, on same scale
- used to compare two or more similar variables

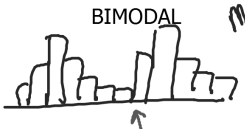
Example: *(on calculator)*
Create parallel boxplots for the lists SATMF and SATMM (SAT math scores for males and females)

DESCRIBING DISTRIBUTIONS: Shape, Center, Spread

SHAPE

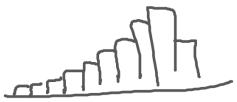
MODE: UNIMODAL BIMODAL multimodal



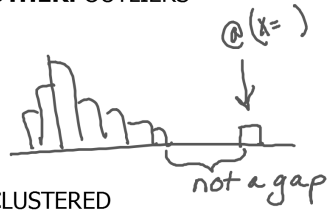


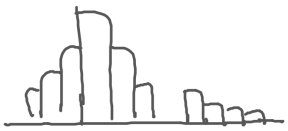
SHAPES: UNIFORM SYMMETRIC

LEFT SKEWED RIGHT SKEWED



OTHER: OUTLIERS GAPS

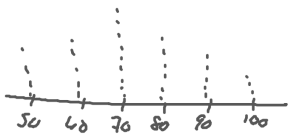




CLUSTERED GRANULARITY

- Consistent, understandable gaps

Ex: MC Quiz
10 ?'s



CENTER:

- Mean or Median
  - * Mean --> Symmetric
  - * Median --> Skewed or outliers

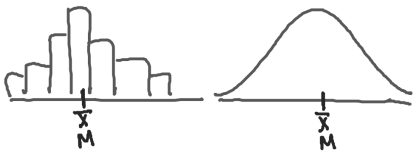
SPREAD:

Goes along with what center you chose

Mean --> std. dev. & range
Median --> IQR & range

Mean vs Median (in distributions)

* Symmetric:
mean ≈ median
300 305
2.8 2.9



* Left skewed:
mean < median

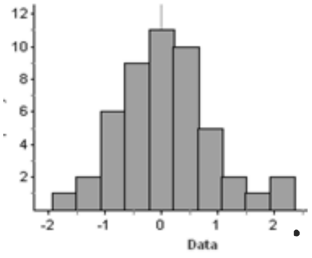
* Right skewed:
mean > median

$\bar{x} = 500$
$M = 502$

symm

$\bar{x} = 26$
$M = 52$

Examples:

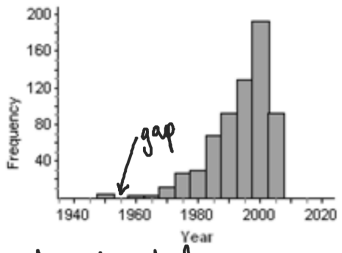


mean = 0.2 x
std dev = 0.677 x
Median = 0
Q1 = -0.7
Q3 = 0.9
min = -1.8
max = 2.4

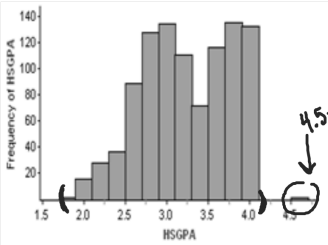
- roughly symmetric, unimodal
- Center @ mean of 0.2 units.
- Std. dev. of 0.677 units.
  range of (-1.8, 2.4) units.

mean = 1982
std dev = 25
Median = 1995 x
Q1 = 1985
Q3 = 2001
min = 1950
max = 2009

IQR



- left skewed, unimodal
- Center @ median of 1995 years
- IQR of 16 years, range of (1950, 2009) years
- Gap @ approx. 1955 yrs.

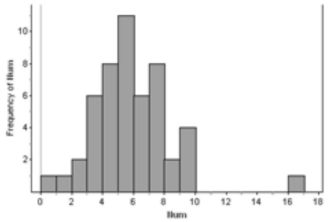


mean = 2.9
std dev = 0.3
Median = 3.35 x
Q1 = 2.7
Q3 = 3.6
min = 1.6
max = ~~4.7~~ 4.70

IQR

- left skewed, bimodal
- Center @ median of 3.35 pts.
- IQR of 0.9 pts, range of (1.6, 4.1) pts.
- Outlier @ the 4.5-4.75 bar.

mean = 7.2
std dev = 2.7
Median = 5
Q1 = 3
Q3 = 7.5
min = 0
max = 16



- roughly symm., poss. slight left skew; unimodal
- Center @ med. of 5 units
- IQR of 4.5 units, range of (0, 16) units
- Outlier @ 16 units

Describing & Comparing two (or more) **distributions:**
- Mention shape, center, and spread of each, and use a comparison word between each

Example: The median for Class 1 was 76 points which ***was higher than*** the median for Class 2, which was 73 points.

When comparing:

symmetric to symmetric	use Mean, s, range
skewed to skewed (outliers)	use Med, IQR, range
skewed to symmetric	use Med, IQR, range

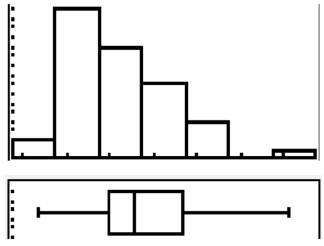
** CANNOT COMPARE MEAN TO MEDIAN**

Example:
Compare and describe the data sets SATMF and SATMM

- The males are right skewed & unimodal which is different than females which are left skewed & unimodal.
- The center for males is Median of 610 pts, which is lower than the female Med. of 700 pts.
- The IQR of males is 55 pts, which is smaller than females of 120 pts.
- The males range is (480, 700) pts which is smaller than females range of (500, 800) pts.

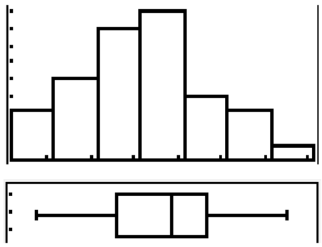
Example 2: compare and describe SMOKERS and NONSMOKERS (from HW problem #32)

smoke:



Summary Stats- SMOKERS
$\bar{x}$ = 237.98
s = 38.54
min = 155
Q1 = 211
Med = 230
Q3 = 267
Max = 351
IQR = 56

non-smokers



Summary Stats- NON-SMOKERS
$\bar{x}$ = 233.06
s = 47.68
min = 134
Q1 = 196
Med = 238
Q3 = 265
Max = 328
IQR = 69

2) The distribution of Smokers is strongly right skewed, whereas the distribution of Non-Smokers is roughly symmetric. Both are unimodal. There is a possible gap around 320 points for smokers.

The center for the Smokers is at the median of 230 points, while the Non-Smokers median is higher at 238 points.

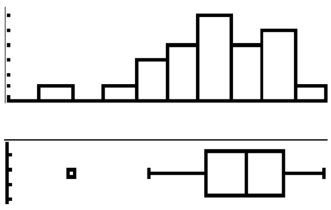
The spread for the Smokers is the IQR of 56 points, which is lower than the IQR of the Non-Smokers at 69. The range of the Smokers is (155, 351) points, which is roughly the same width as the Non-Smokers range of (134, 328) points.

Complete Ch. 4 & 5 practice worksheet

except #6

BOXPLOT EXPLORATION

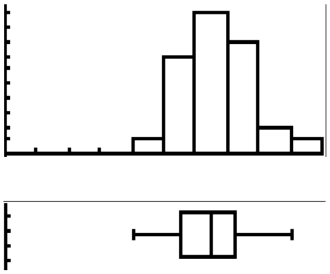
BLK1



min = 4
Q1 = 12.5
Med = 15
Q3 = 17.5
Max = 20

mean = 14.48
std.dev = 3.676

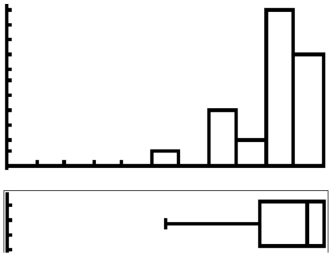
BLK2



min = 8
Q1 = 11
Med = 13
Q3 = 14.5
Max = 18

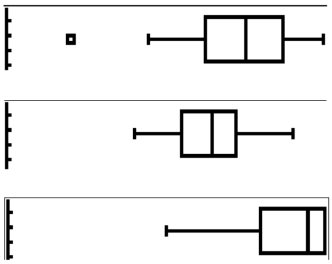
mean = 12.93
std.dev = 2.154

BLK3



min = 10
Q1 = 16
Med = 19
Q3 = 20
Max = 20

mean = 17.923
std.dev = 2.481



Describe: